doi:10.3772/j.issn.1002-0470.2013.05.002

GATS-LSVM:新的网络入侵检测方法①

李文法②*** 孙连英*** 刘 畅*** 马小军***

(*北京联合大学信息服务工程重点实验室 北京 100101)

(**北京联合大学信息学院 北京 100101)

摘 要 针对现有网络入侵检测方法的不足,提出了一种新的网络入侵检测方法——GATS-LSVM 算法。该方法采用遗传算法(GA)与禁忌搜索(TS)相混合的搜索策略对特征子集空间进行随机搜索,利用提供的数据在无约束优化线性支持向量机(LSVM)上的分类错误率作为特征子集的评估标准获取最优特征子集,从而有效地对入侵进行检测。大量基于著名的 KDD Cup 1999 数据集的实验表明,该新方法相对于其它一些传统的网络入侵检测方法,能在保证较高检测率的前提下,有效地降低误报率、入侵检测的计算复杂度和提高检测速度,能更适用于现实高速网络应用环境。

关键词 网络入侵检测,遗传算法(GA),禁忌搜索(TS),线性支持向量机(LSVM)

0 引 言

当前网络安全问题日益突出,如何迅速、有效地发现各类新的入侵行为,对于保证网络安全显得十分重要。网络入侵检测系统是网络安全防御体系的一个重要组成部分[1]。入侵检测系统通过对网络和主机上某些关键信息进行收集分析,检测是否有违反安全策略的事件或攻击事件发生,若已发生,便发出警报。入侵检测系统处理的数据中含有大量相关与冗余信息,影响了入侵检测的准确率和速度,因而对这些数据进行修剪与剔除显得尤为重要。特征选择是网络入侵检测系统的一个重要组成部分,它针对网络数据流的高维特征空间存在大量的相关与冗余特征的特性,在此高维空间上应用搜索算法来寻找最优的特征子集,剔除那些冗余的特征。在得到最重要和必要的特征形成向量后再进行训练和检测,这不仅可以降低入侵检测系统的计算复杂性和提高检测速度,而且可以获得很好的检测效果。特征选择需要有评估特征子集的标准和搜索最优特征子集的策略。针对大数据集,在评估标准方面,文献[2]将支持向量机(SVM)引入特征选择,并给出了相关算法,分类信息明确的特征可以被该算法选出;文献[3]使用向量机和递归属性排除作属性评估。但这些算法的评估速度并不理想。在搜索策略方面,文献[4]提出了启发式搜索策略,例如顺序搜索、反向搜索、正向搜索等;文献[5]提出了随机搜索策略,如遗传算法,相比于启发式搜索具有一定的优势。但这些搜索策略收敛速度慢,计算资源耗用大,常常得到局部最优解,进行特征选择时并不适用于大数据集的入侵检测。文献[6-8]分别提出了基于遗传算法(GA)和 SVM(GA-SVM)的网络入侵检测方法(简称 GA-SVM 算法)、基于相关性选择(correlation feature selection, CFS)和 SVM(CFS-SVM)的网络入侵检测方法(简称 CFS-SVM 算法)、基于主成分分析(PCA)和 C4.5 决策树(PCA-C4.5)的网络入侵检测方法(简称 PCA-C4.5 算法),但这些方法在特征选择上花费较多的时间,入侵检测效果并不令人十分满意。针对上述缺点,本文提出了一种基于遗传算法(GA)与禁忌搜索(tabu search, TS)(GATS)和线性支持向量机(LSVM)的网络入侵检测方法,简称 GATS-LSVM 算法。实验表明,相对于传统的网络入侵检测方法,该方法在保证较高检测率的前提下,可有效降低误报率,减少入侵检测的计算复杂度,提高检测速度,更适用于现实高速网络应用环境。

① 863 计划(2007AA01Z416),北京联合大学新起点计划(ZK201204)和人才强校计划人才(BPHR2011A04)资助项目。

② 男,1974 年生,博士,副教授;研究方向:网络安全,入侵检测,数据挖掘;联系人,E-mail: liwenfa@buu.edu.cn

(收稿日期:2012-11-06)

1 GATS-LSVM 算法

本文提出的网络入侵检测算法包括搜索策略——遗传算法与禁忌搜索相混合(GATS)策略和评估标准——线性支持向量机(LSVM)评估两部分。

1.1 GATS 搜索策略

遗传算法(GA)已广泛用于求解函数优化问题,是一种进化搜索算法[9]。遗传算法的两个重要组成部分是重组和变异算子。禁忌搜索(TS)是由 Glover 提出的启发式搜索算法[10],禁忌搜索独特的特点之一是具有记忆功能。Glover 从理论上分析和论述了混合 GA 与 TS 的必要性和可行性[11]。为了保持原算法的优点,克服或弱化不足,进而提高算法的力度,在上述理论的基础上,本文将 GA 与 TS 进行了混合,构造了一种新的混合搜索策略 GATS。在 GA 进化搜索过程中引入 TS 独有的记忆功能,给出了新的禁忌搜索重组(tabu search recombination, TSR)算子。利用 TS 的优点(爬山能力强)去克服 GA 的缺陷(爬山能力差),即用 TS 取代 GA 的禁忌搜索变异(tabu search mutation, TSM)算子去增强 GA 的爬山能力。这样,GATS 策略就保持了 GA 的多出发点的优势,并具有了 TS 的爬山能力强和记忆功能。后面的实验验证了混合算法 GATS 的优势。

禁忌搜索重组(TSR)算子使用一个禁忌表(长度为 T)来记录染色体的适应值,父代群体适应值的平均值等于渴望水平。在进行禁忌搜索重组时,比较子代的适应值和渴望水平,如果子代适应值好于渴望水平,则这个染色体被禁,并进入到下一代中。如果子代适应值差于渴望水平,且属于禁忌,则将最好的那个父代放入到下一代中。图 1 描述了 TSR 过程。在 TSR 过程中,进入到下一代的是大部分具有高适应值的子代。TSR 通过禁忌表来限制次数,适应值相同的子代就不会出现太多,因此染色体结构的多样性在群体中就尽可能被保持,算法不容易早熟。

```
Begin
    if fitness of x > average value of population
      then accept x;
    else
      if offspring x is not in tabu list
        accept x
      else
          choose the better of two parents to the next generation;
      update tabu list;
end
```

图 1 TSR 过程伪代码

在操作禁忌搜索变异(TSM)算子时,首先把一个染色体作为 TSM 的输入初始解,通过禁忌搜索变异,输出一个解。与标准变异算子不同之处在于,在 TSM 的变异搜索过程中,移动值需要调用评价函数去确定,通过比较禁忌表 T 和移动值来决定输出哪个移动值。另外,在禁忌搜索变异过程中也可以接受劣解,因此 TSM 具有爬山能力。假设 x 代表一个染色体,图 2 描述了 TSM 的操作过程。

```
Begin
    t=0; set the best solution x(0)=x; set T;
    while termination condition not satisfied do
        t=t+1;
        move x to x';
        update(x, x(0), tabu list);
end
```

图 2 TSM 过程伪代码

1.2 无约束优化线性支持向量机

支持向量机[12,13](SVM)以统计学习理论为基础,能够较好地解决非线性、小样本、高维数和局部最小点等实际问题。图 3 显示出了其思想。在图 3 中,空心点和实心点表示两类不同的样本,H 为分类超平面,H_1,H_2 分别代表各类中离 H 最近的样本且平行于 H 的面,它们之间的距离称为分割距离。所谓最优分类面就是要求不但能将两类正确分开,而且使分类间隔最大。H_1,H_2 上的样本点称为支持向量。

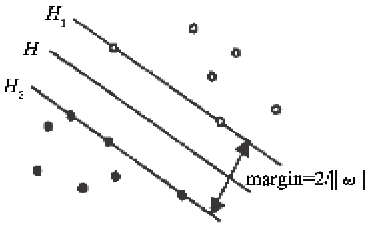

图 3 线性可分情况下的最优分类面

对于训练样本为 $\{(x_i,y_i)\}_{i=1}^{l} \subset \boldsymbol{R}^n \times \{1,-1\}$ 的二分类问题,根据统计学理论,可建立如下标准的线性 SVM 模型 A:

$$\begin{cases} \min \dfrac{1}{2}(\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{\omega}) + C\sum\limits_{i=1}^{l}\xi_i \\ \text{s.t. } yi((\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{x}_i) + b) \geqslant 1 - \xi_i,\ i = 1,2,\cdots,l \\ \xi_i \geqslant 0, i = 1,2,\cdots,l \end{cases} \tag{1}$$

其中 $C>0$ 为正则化参数，$\xi_i(i=1,2,\cdots,l)$ 为松弛变量，$\boldsymbol{\omega}\in\boldsymbol{R}^n$ 为分类超平面的法向量，$b\in\boldsymbol{R}$ 为阈值。利用优化理论中的 KKT 条件和对偶理论，可得对偶优化模型 A′

$$\begin{cases}\max \sum_{i=1}^{l}\alpha_i-\frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}\alpha_i\alpha_j y_i y_j(x_i^{\mathrm{T}}x_j)\\ \text{s.t.}\ \sum_{i=1}^{l}y_i\alpha_i=0\\ 0\leqslant\alpha_i\leqslant C,\ i=1,2,\cdots,l\end{cases}\tag{2}$$

其中 $\alpha_i(i=1,2,\cdots,l)$ 为 lagrange 乘子。优化问题 A′是一个凸二次规划问题，其局部最优解即为全局最优解。若 $\boldsymbol{\alpha}^*=(\alpha_1^*,\alpha_2^*,\cdots,\alpha_l^*)^{\mathrm{T}}$ 为模型 A′的最优解，则

$$\boldsymbol{\omega}^*=\sum_{i=1}^{l}\alpha_i^*y_ix_i\tag{3}$$

根据 KKT 互补条件，最优解必满足

$$\alpha_i(y_i(\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{x}_i+b)-1+\xi_i)=0\quad i=1,2,\cdots,l\tag{4}$$

$$(\boldsymbol{C}-\alpha_i)\xi_i=0,\ i=1,2,\cdots,l\tag{5}$$

由式(3)－(5)可知，对应于 lagrange 乘子 $\alpha_i=0$ 的样本对分类问题不起什么作用，而只有对应于 lagrange 乘子 $\alpha_i>0$ 的样本(支持向量)对计算 $\boldsymbol{\omega}^*$ 起作用，从而决定分类结果，支持向量通常只是全体样本中的很少一部分。求解上述问题后得到的广义最优线性分类器为

$$f(\boldsymbol{x})=sgn\{(\boldsymbol{\omega}^{*\mathrm{T}}\boldsymbol{x})+b^*\}=sgn\{\sum_{i=1}^{l}\alpha_i^*y_i(\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{x})+b^*\}\tag{6}$$

其中 $sgn(\cdot)$ 为符号函数，b^* 为分类的阈值，可通过任意一个支持向量求得。模型 A 与 A′都是约束条件下的优化模型，为了得到线性支持向量机的无约束优化模型，不妨定义

$$g_i(\boldsymbol{\omega},b)\triangleq 1-y_i(\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{x}_i+b)\quad i=1,2,\cdots,l\tag{7}$$

则由式(4)、(5)可得

$$\boldsymbol{\xi}_i=\max\{0,g_i(\boldsymbol{\omega},b)\},\ i=1,2,\cdots,l\tag{8}$$

将其代入标准的线性 SVM 模型 A 中，可得线性 SVM 的无约束优化模型 B

$$\min\frac{1}{2}\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{\omega}+C\sum_{i=1}^{l}\max\{0,\ g_i(\boldsymbol{\omega},\ b)\}\tag{9}$$

由最优化理论知，模型 A 与模型 B 等价，模型 B 目标函数中的前两项恰好体现了统计学习理论中的结构风险最小化原则。其中前一项反映了模型的置信范围，后一项反映了模型的训练误差。注意到

$$\|\boldsymbol{\xi}\|_1=\sum_{i=1}^{l}\max(0,\ g_i(\boldsymbol{\omega},b))\tag{10}$$

$$\|\boldsymbol{\xi}\|_\infty=\max\{0,\ g_1(\boldsymbol{\omega},b),\cdots,g_l(\boldsymbol{\omega},b)\}\tag{11}$$

如果用向量 $\boldsymbol{\xi}=(\xi_1,\xi_2,\cdots,\xi_l)^{\mathrm{T}}$ 的 ∞ 范数来度量模型的训练误差，则可得无约束优化模型 C

$$\min\Phi(\boldsymbol{\omega},b)=\frac{1}{2}\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{\omega}+C\max\{0,g_1(\boldsymbol{\omega},b),\cdots,g_l(\boldsymbol{\omega},b)\}\tag{12}$$

与 A、A′相比，SVM 无约束优化模型 B、C 在数学形式上更加简洁明了。但由于优化问题 B、C 是一个和最大值函数有关的一类不可微优化问题，从而给求解带来了困难。一条可行的途径是利用光滑化技术将不可微优化问题转化为可微优化问题，从而易于求解。本文利用极大熵方法作为求解优化问题 C 的一种近似解法。极大熵方法基本思想是[14,15]：对于极大极小问题 $\min\phi(\boldsymbol{x})=\max\{f_1(\boldsymbol{x}),f_2(\boldsymbol{x}),\cdots,f_m(\boldsymbol{x})\}$，利用最大熵原理推导出一个可微函数 $\phi_p(\boldsymbol{x}=\frac{1}{p}\ln(\sum_{i=1}^{m}\exp(pf_i(\boldsymbol{x}))))$，通常称为极大熵函数。用该可微函数来逼近最大值函数 $\phi(\boldsymbol{x})$，从而把不可微优化问题转化为可微优化问题，使问题简化。通过引入极大熵函数，问题 C 的求解被转化为如下的无约束优化问题 D：

$$\min\Phi_p(\boldsymbol{\omega},b)=\frac{1}{2}\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{\omega}+\frac{C}{p}\ln(1+\sum_{i=1}^{l}\exp(pg_i(\boldsymbol{\omega},b)))\tag{13}$$

其中 $p>0$ 是参数。由极大熵函数的逼近性质[16]和得出的相关定理[17,18]可以得出，问题 C 和 D 是等价的，并且优化问题 D 的任一局部最优解都是全局最优解。下面给出标准无约束优化算法的子程序。

基本算法：

步骤 1 给定任一初始点 $(\boldsymbol{\omega}^{(0)},\ b^{(0)})$ 和正则化参数 C，令 p 为一个充分大的常数。

步骤 2 用无约束优化算法子程序进行 $\Phi_p(\boldsymbol{\omega},b)$ 的最小化计算。

步骤 3 将最优解 $(\boldsymbol{\omega}^*,\ b^*)$ 代入式(6)，从而得到广义最优线性分类器。

2 基于 GATS-LSVM 的入侵检测

基于 GATS-LSVM 的网络入侵检测框架见图 4。

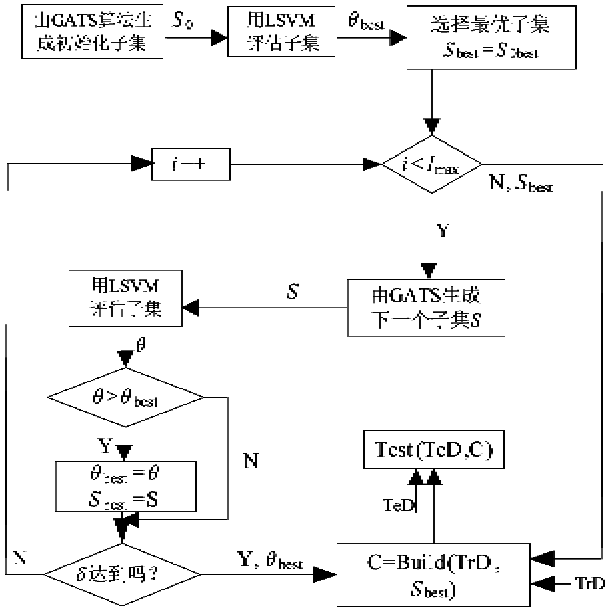

图 4　基于 GATS-LSVM 的入侵检测框架

首先由 GATS 算法生成一个随机特征子集 S_0，并输入作为初始子集，S_0 经过优化的线性支持向量机评估，得出最优评估结果 θ_{best}。初始化时，变量 S_{best} 的初始值为 S_{0best}，它的初始评估值为 θ_{best}。在迭代过程中，每一次迭代，均通过优化的线性支持向量机来评估特征子集集合 S，产生最优评估值 θ，S_{best} 为相应的特征子集。若 θ_{best} 小于 θ，则 θ 的值赋值给 θ_{best}，S_{best} 等于 S 的值。若 θ 值大于 θ_{best} 值，则继续下一次迭代。在迭代的最后阶段，判断是否满足了预先设定的标准 δ，标准 δ 是根据实际需要，综合衡量而给出的一个经验值，若迭代满足了要求的标准，则结束迭代，并把生成的最优特征子集 S_{best} 输送给分类器，用于对应的入侵检测。若没有满足预先设定的标准 δ，且没有达到最大的迭代次数，则继续下一轮迭代。选择好特征后，利用选择的特征子集在训练集 TrD 上训练模型 C，最后通过测试集 TeD 测试模型 C 的性能。由于应用 GATS-LSVM 算法之后，特征空间下降，特征空间中相关与杂音特征被剔除，基于选择后特征的入侵检测系统的结构变得简洁清晰，检测速度应该会提高很多，入侵检测系统的检测率也有一定程度的提高。

3　实验研究

实验的目的是验证本文提出的 GATS-LSVM 算法的网络入侵检测的有效性。实验主要包括：与效果比较好的基于遗传算法和支持向量机(GA-SVM)的网络入侵检测方法、基于相关性选择和支持向量机(CFS-SVM)的网络入侵检测方法、基于主成分分析和 C4.5 决策树(PCA-C4.5)的网络入侵检测方法的检测率和误报率进行比较；对比使用 GATS-LSVM 选择的特征和使用全部 41 种特征建立检测模型的接收者操作特性(receiver operating characteristic, ROC)曲线的分值；从建模时间和检测时间两方面对比 GATS-LSVM 方法与其它三种检测方法的检测速度。

3.1　数据集及实验环境

为了保证实验的说服力和方便性，我们采用研究领域共同认可及广泛使用的基准评测数据集 KDD Cup 1999[19] 进行测试。KDD Cup 1999 数据集是关于入侵检测的一个标准数据集，主要分为训练数据集和测试数据集两部分。该数据集包括大约 490 万条数据记录，每条都是从网络环境中模拟攻击所得的原始网络数据中根据设定的 41 个特征提取出来的，它们都是描述网络连接统计信息的特征向量，包含 5 类数据：DoS、Probe、R2L、U2R 四类攻击数据(共包含 24 种攻击类型)以及正常数据。本文所述的实验都在同一平台下完成，该平台的配置为：Intel processor 2.0GHz，2048MBRAM，Windows 7 操作系统。

3.2　真阳性率和假阳性率的对比

在实验开始之前，为满足实验的需要，我们对 KDD Cup 1999 数据集进行了一些预处理工作：将 KDD Cup 1999 的数据集进行了提取；在数据集中随机提取了 1986 条正常数据(统一标记为“0”)和 2056 条攻击数据(包括上述 4 类攻击，统一标记为“1”)，其中攻击数据占整个数据集的 1.2%。我们采用了 10 折交叉验证(ten fold cross validation)的方法，重复实验 10 次，取真阳性率(true positive rate, TPR)和假阳性率(false positive rate, FPR)的平均值对几种算法进行了对比。其中，TPR 定义为正确检测出的攻击样本的数量/总的攻击样本的数量，FPR 定义为错误判为攻击的正常样本数量/总的正常样本的数量。

在实验中，我们利用 WEKA[20] 软件并挑选不同参数对 4 种网络入侵检测方法即 GA-SVM 算法、CFS-SVM 算法、PCA-C4.5算法和本文提出的 GATS-LSVM 算法进行了多次实验，各取其检测效果最好的结果作为比较。实验结果如表 1 所示。从表中我们不难看出，在训练数据充足的情况下，GATS-LSVM 算法的真阳性率要优于其它三种方法，而且假

阳性率也相对较低。

表 1 四种不同的检测方法十字交叉对比实验结果

检测方法	TPR(%)	FPR(%)
GA-SVM	97.6	1.85
CFS-SVM	97.8	2.33
PCA-C4.5	98.8	0.83
GATS-LSVM	99.2	0.80

同时,为了验证使用 GATS-LSVM 算法的入侵检测比没有使用 GATS-LSVM 算法的入侵检测在检测已知攻击和未知攻击上有更高的真阳性率,我们也进行了相应的实验,比较结果如图 5 所示。在图 5 中,我们把所有攻击看作一种类型,并建立两种系统:使用所有特征的入侵检测系统和使用 GATS-LSVM 算法的入侵检测系统。针对每种系统,我们使用两种测试集进行测试,一种是已知攻击(取样的训练集中存在的攻击)测试集,一种是未知攻击(取样的训练集中不存在的攻击)测试集。从图 5 可以看出,在检测已知攻击和未知攻击时,与使用所有特征的入侵检测系统相比,使用 GATS-LSVM 算法的入侵检测系统有更高的 ROC 曲线分值,特别是在检测未知攻击时,ROC 曲线分值更高。

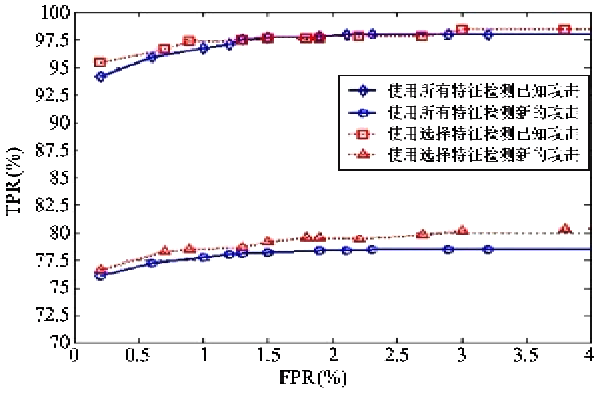

图 5 攻击检测的 ROC 曲线

3.3 检测速度的对比

在训练集上分别使用 GA-SVM、CFS-SVM 、PCA-C4.5 和 GATS-LSVM 四种不同的算法选择最优特征子集。选出最优特征子集之后,分别在 41 个特征和选出的特征子集上建立入侵检测模型。针对训练数据集,使用不同的阈值,建立多个入侵检测模型,综合评价使用 GA-SVM、CFS-SVM 、PCA-C4.5 和 GATS-LSVM 四种算法选择出的特征子集建立的入侵检测模型和使用所有特征子集建立的入侵检测模型在建模时间、检测时间上的不同。入侵检测模型的平均建模时间和检测时间如表 2 所示。从表 2 可以看出:与使用其它几种算法建立的入侵检测模型相比,使用 GATS-LSVM 算法建立的入侵检测模型有更少的建模时间和检测时间,检测速度要好于其它几种算法建立的入侵检测模型;与使用所有特征子集的入侵检测模型相比,使用 GATS-LSVM 算法选择特征子集的入侵检测模型有更少的建模时间和检测时间。例如,对于使用 GATS-LSVM 算法的入侵检测模型,使用所有特征子集的入侵检测模型平均建模时间和检测时间分别是 79s 和 21s,而使用 GATS-LSVM 算法的选择特征子集的入侵检测模型相应的时间分别是 31s 和 8s,分别是使用所有特征子集的入侵检测模型的39.2% 和38.1% 左右。根据建模时间和检测时间的比较结果,可以看到,与 GA-SVM 算法、CFS-SVM 算法相比,我们提出的 GATS-LSVM 算法消耗更少的计算资源,检测速度更快。同时,使用 GATS-LSVM 算法选择特征子集的模型有更少的建模时间和检测时间,说明 GATS-LSVM 算法提高了入侵检测的速度。

表 2 基于四种不同算法的检测模型在所有特征和选择特征上的平均建模时间和检测时间

—		GA-SVM	CFS-SVM	PCA-C4.5	GATS-LSVM
建模时间(s)	所有	121	152	83	79
	选出	61	65	35	31
检测时间(s)	所有	31	46	26	21
	选出	11	16	9	8

4 结 论

本文提出了一种网络入侵检测新方法——GATS-LSVM 算法,该方法采用遗传算法(GA)与禁忌搜索(TS)相混合的搜索策略对特征子集空间进行随机搜索,利用提供的数据在无约束优化线性支持向量机(LSVM)上的分类错误率作为特征子集的评估标准获取最优特征子集,从而有效地对入侵进行检测。大量基于著名的 KDD Cup 1999 数据集的实验表明:相对于其它一些传统的网络入侵检测方法,GATS-LSVM 算法在保证较高检测率的前提下,能有效地降低了误报率和较大的训练集规模给入侵检测方法带来的过大计算开销,极大地提升其在现实网络环境中的可用性。

本文提出的网络入侵检测方法还需要进一步地

改进,并提高其性能,以满足实际网络环境应用的需要:在实际的网络环境中,如何应用和优化,在现实的大流量网络中,如何检测分布式拒绝服务攻击和检测 Web 服务器的异常情况,它们将是我们下一步的工作重点。

参考文献

[1] Zaman S, Karray F. Lightweight IDS based on features selection and IDS classification scheme. In: Proceedings of the 12th IEEE International Conference on Computational Science and Engineering. Vancouver, Canada, 2009. 365-370

[2] Grandvalet Y, Canu S. Adaptive scaling for feature selection in SVMs. *Advances in Neural Information Processing Systems*, 2003,15:553-560

[3] Mao K. Feature subset selection for support vector machines through discriminative function pruning analysis. *IEEE Transactions on Systems, Man, and Cybernetics*, 2004,34(1):60-67

[4] Jain A, Zongker D. Feature selection: Evaluation, application, and small sample performance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1997, 19(2): 153-158

[5] Kudo M, Sklansky J. Comparison of algorithms that select features for pattern classifiers. *Pattern Recognition*, 2000, 33(1): 25-41

[6] Kim D, Nguyen H N, Ohn S Y. Fusions of GA and SVM for anomaly detection in intrusion detection system. In: Proceedings of the 2nd IEEE International Symposium on Neural Networks, Chongqing, China, 2005. 415-420

[7] Shazzad K M, Jong S P. Optimization of intrusion detection through fast hybrid feature selection. In: Proceedings of the 6th International Conference on Parallel and Distributed Computing, Applications and Technologies, Dalian, China, 2005. 264-267

[8] Chen Y, Dai L, Li Y, et al. Building efficient intrusion detection model based on principal component analysis and C4.5 algorithm. In: Proceedings of the 9th IEEE International Conference on Advanced Communication Technology, Phoenix Park, Korea, 2007. 2109-2112

[9] Holland J. Adaptation in Natural and Artificial Systems. Cambridge: The University of Michigan Press, 1975. 65-78

[10] Glover F. Future paths for integer programming and links to artificial intelligence. *Computers and Operations Research*, 1986,13(4):533-549

[11] Glover F, Kelly J, Laguna M. Genetic algorithms and tabu search:hybrids for optimizations. *Computers and Operations Research*, 1995, 22(1):111-134

[12] 陈友,程学旗,李洋等. 基于特征选择的轻量级入侵检测系统. 计算机学报,2007, 18(7):1639-1651

[13] Vapnik V. The Nature of Statistical Learning Theory. New York: Springer Verlag, 1995. 17-23

[14] 李兴斯. 非线性极大极小问题的一个有效解法. 科学通报,1991,36(19):1448-1451

[15] 李兴斯. 一类不可微优化问题的有效解法. 中国科学(A 辑),1994,24(4):371-377

[16] 唐焕文,张立卫,王雪华. 一类约束不可微优化问题的极大熵方法. 计算数学, 1993,15(3):268-275

[17] 唐焕文,张立卫. 凸规划的极大熵方法. 科学通报, 1994,39(8):682-684

[18] 张志华,郑南宁,史罡. 极大熵聚类算法及其全局收敛性分析. 中国科学(E 辑),2001,31(1):50-70

[19] KDD cup 1999 data. http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html

[20] Nguyen T T, Duong T A. Comparing three improved variants of simulated annealing for optimizing dorm room assignments. In: Proceedings of IEEE International Conference on Computing and Communication Technology, Danang, Vietnam, 2009. 1-5

GATS-LSVM: a new network intrusion detection method

Li Wenfa[*,**], Sun Lianying[*,**], Liu Chang[*,**], Ma Xiaojun[*,**]

([*] Beijing Key Laboratory of Information Service Engineering, Beijing Union University, Beijing 100101)

([**] College of Information Technology, Beijing Union University, Beijing 100101)

Abstract

Aiming at the shortcomings of existing network intrusion detection approaches, a new network intrusion detection scheme, called the GATS-LSVM algorithm, was proposed. The scheme uses the hybrid genetic algorithm (GA) and tabu search (TS) as the search strategy to specify a candidate subset for evaluation, and then uses the linear support vector machine (LSVM) as the evaluation function to obtain the optimum feature subset by the data classification error rate, so as to boost the detection performance. A series results of the experiment conducted based on the KDD cup 1999 dataset demonstrate that the proposed method has the higher detection rate and lower false alarm rate. Furthermore, the proposed method can reduce computational resources of intrusion detection, improve the detection speed and is more suitable for the real network applications than the traditional ones.

key words: network intrusion detection, genetic algorithm(GA), tabu search(TS), linear support vector machine(LSVM)